Post-transcriptional Regulation of *Hes7* via Alternative Polyadenylation of the 3'UTR

Undergraduate Research Thesis

Presented in partial fulfillment of the requirements for graduation *with honors research distinction* in the undergraduate colleges of The Ohio State University

By

Madeline Parker

The Ohio State University
May 2016

Project Advisor: Dr. Susan E. Cole, Department of Molecular Genetics

ABSTRACT

During vertebrate embryonic development, the axial skeleton forms from transient structures called somites during the process of somitogenesis. This process is regulated by a segmentation clock which allows embryos to integrate temporal and positional information through synchronized pulses of gene expression. The period of the clock changes across the presomitic mesoderm (PSM) such that the cells in the anterior PSM experiences slower oscillations than those in the posterior PSM. Currently, the mechanism by which the clock period changes across the PSM is unclear. Additionally, the clock period varies between species but is consistent within a species. Many models suggest that the oscillation rate of the clock in mice is set by the period of *Hes7* oscillations. If the speed of *Hes7* oscillations determines the period of the clock, then mechanisms that regulate *Hes7* transcript turnover may be important for normal clock function and may influence the clock period. This is because the rate of transcript turnover, determined by transcript half-life, affects the amount of time that lapses between turning off transcription and turning off gene activity by degrading all mRNA and protein. Regulation of transcript stability and turnover is often controlled by the 3' untranslated region (3'UTR). In *Hes7,* there are three 3'UTR isoforms of different lengths produced through alternative polyadenylation. This means that the use of alternative polyadenylation sites may influence *Hes7* turnover in the PSM and contribute to the change in clock frequency across the PSM or between species. To test whether the different *Hes7* 3'UTR isoforms could impact RNA turnover, the *Hes7* 3'UTRs have been mutated to force the expression of a single isoform. These forced-polyadenylation UTRs were tested to analyze their impact on transcript stability through Luciferase assays. We find that the *Hes7* 3'UTR is associated with decreased protein production compared to controls and that transcripts containing the longer 3'UTR isoforms produce

significantly less protein than the short isoforms. Further research will determine whether the functional differences between 3'UTR isoforms arise due to changes in transcript turnover or changes in translational efficiency and will examine mechanisms regulating these effects.

CHAPTER 1: BACKGROUND

Somitogenesis

During vertebrate embryonic development, the body is segmented in a process called somitogenesis. The precursors to vertebrae, ribs, and the surrounding soft tissue are first formed as transient embryonic structures called somites. Proper formation of somites is necessary for the correct formation of the axial skeleton, which includes the vertebrae and ribs, and perturbations to somitogenesis can cause skeletal defects (reviewed in Wahi 2016 and Pourquie 2011). Because of the potential for skeletal deformations, it is important to understand the regulation of somitogenesis. Proper vertebrae and rib formation is necessary for the spinal joints to form correctly to allow normal movement. In humans, improper segmentation can result in several congenital defects including congenital scoliosis, spondylocostal dysostosis, and spondylothoracic dysostosis (Penton 2012; Pourquie 2011). Several of these defects have been linked to mutations in genes associated with the Notch signaling pathway (Penton 2012; reviewed in Pourquie 2011).

Somitogenesis begins with the formation of an unsegmented growth zone called the presomitic mesoderm (PSM) at the posterior end of the embryo. Cells enter the PSM from the blastopore during early segmentation and from the tailbud during later segmentation (reviewed in Wahi 2016). Somites form sequentially along the anterior-posterior axis by budding from the anterior PSM in pairs beginning with the anterior-most somites (reviewed in Wahi 2016). While the number of somites formed, and rate of formation, is consistent within embryos of a specific species, it varies between species. On average, somites are formed every 2 hours in mice, 90 minutes in chicks, 30 minutes in zebrafish, and 4-6 hours in humans (reviewed in Wahi 2016).

The clock and regulation of somitogenesis

The PSM has two regions: the posterior PSM (region I) and the anterior PSM (region II) (Fig. 1A). In somitogenesis, cells enter region I of the PSM and experience synchronized pulses of clock-gene expression. Clock oscillations slow as cells move anteriorly, displaced by new cells entering the PSM, and oscillations stop altogether in region II. In region II, somite-sized clusters of cells are patterned to define caudal and rostral compartments. These compartments, distinguishable by different gene expression patterns, will ultimately contribute to different tissues that arise from somites (as reviewed in Wahi 2016). Boundaries between somites are formed via a mesenchymal-to-epithelial transition that creates a cleft between the new somite and the anterior PSM, bordered on each side by epithelial cells (Pourquie 2011).

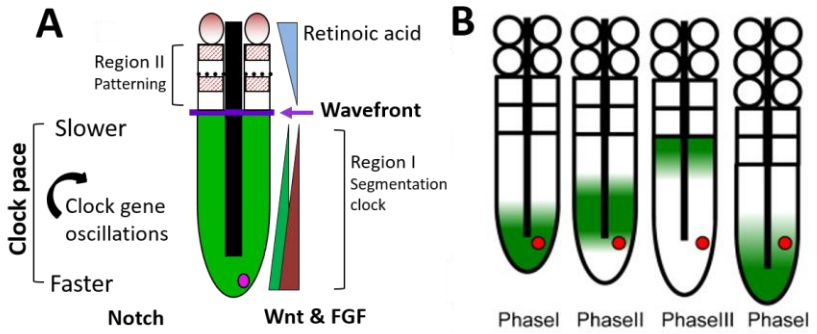

Figure 1. PSM and clock gene expression
A) Schematic representation of the PSM. Clock genes oscillate in region I. The wavefront separates region I from region II, where oscillations are halted.
B) Clock genes exhibit oscillatory expression (green) where the individual red cell experiences pulses of gene expression. Image from Wahi 2016.

Spatial and temporal information must be coordinated in order to create the proper number, size, and arrangement of somites. The clock and wavefront model has been proposed to explain regulation of somitogenesis (Cooke 1976). In this model, a molecular clock synchronizes oscillatory gene expression in the PSM to set the rate of somite formation. Gene expression of the members of the Notch signaling pathway is seen as a stripe of expression that moves anteriorly through the PSM in three distinct phases (Fig 1B). However, individual cells only experience periods of clock gene activation ('on') and inactivation ('off'). One on/off cycle

correlates with the formation of one pair of somites. A wavefront created by Wnt and FGF expression gradients provides spatial regulation to specify the location of intersomitic boundaries. Coordination of these two processes ensures the production of somites of the proper size and spacing (as reviewed in Wahi 2016). It is still unknown why the clock slows anteriorly or what causes the change in period of the clock. Additionally, there is still debate over which genes set the clock period.

The first evidence of a genetic clock was the discovery that expression of the gene *c-hairy1*, an effector of the Notch signaling pathway, oscillates in the PSM of chicken embryos. This observation was followed by additional evidence of a segmentation clock by finding other genes with cyclic expression linked to the Notch pathway in other vertebrates (as reviewed in Wahi 2016). Most studies suggest that there is a single cell-autonomous oscillator in the PSM that is synchronized between cells via cell-cell contact. In mice, perturbation of Notch signaling causes severe skeletal deformations and loss of oscillations of downstream oscillatory genes (Ferjentsik 2009). It is believed that the Notch signaling pathway in mice is critical for both synchronizing the clock and for setting the pace of oscillations.

The Notch signaling pathway

The Notch signaling pathway is utilized in cell-cell signaling (as reviewed in Wahi 2016). Notch receptors are transmembrane proteins on the surface of signal-receiving cells (Fig. 2). The receptor will bind to a Delta/Serrate/Lag-2 (DSL) ligand on a signal-sending cell. This interaction causes the Notch intracellular domain (NICD) to be cleaved, released into the

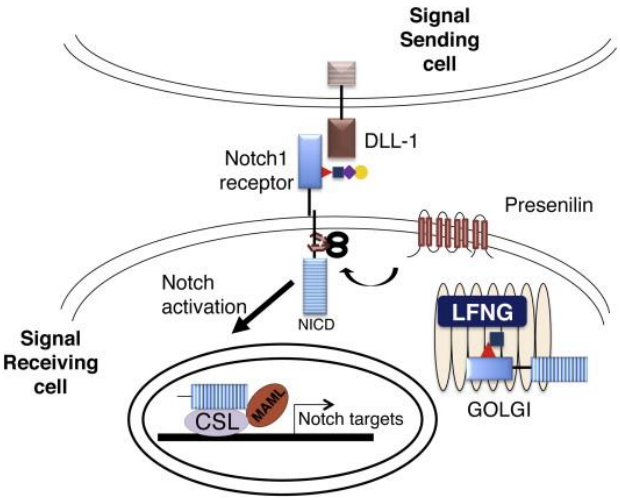

Figure 2. Notch Signaling
The Notch receptor is modified in the Golgi by LFNG and is expressed on the cell surface. Upon activation, the NICD is translocated to the nucleus to drive expression of downstream Notch targets. Figure after Wahi 2016.

cytoplasm of the cell, and translocated to the nucleus to drive expression of downstream Notch targets. Glycosylation of the Notch receptor by Fringe proteins in the Golgi modulates interactions between the Notch receptor and ligands (Fig. 2). Mammalian Lunatic fringe (*Lfng*) has been shown to exhibit oscillatory expression in the PSM and loss of *Lfng* or sustained expression of *Lfng* results in somite fusion, disorganization, and loss of clock oscillations (as reviewed in Wahi 2016). This suggests that *Lfng* oscillations are necessary for somitogenesis, especially in mice, and its regulation of Notch has been proposed as a mechanism to maintain steady Notch oscillations.

It is believed that in mice Notch serves both to promote and to synchronize oscillations of downstream Notch targets such as *Hes7* and Lunatic fringe between cells in the PSM. Many of these downstream target genes are other segmentation clock genes. Segmentation defects are seen in zebrafish, mice, and chickens in which Notch pathway components have been deleted, supporting the idea that Notch is critical for clock gene oscillations (reviewed in Wahi 2016 and Pourquie 2011). In chickens and mammals, there is evidence that Notch may be necessary to produce the oscillations as well as synchronize them (Ferjentsik 2009).

Notch signaling and Hes/her gene expression in the segmentation clock

In all vertebrates examined to date, at least one homologue of the Notch effector hairy/enhancer of split (*Hes/her*) has been shown to oscillate in the PSM (Wahi 2016). In mice, *Hes7* encodes a basic helix-loop-helix type transcription factor which acts as a Notch effector. *Hes7* has been found to oscillate synchronously with *Lfng* in the PSM with a two-hour periodicity in mice. Loss of *Hes7* results in severe skeletal deformations including fused and shortened vertebrae, loss of 4-5 ribs, bifurcated and fused ribs, and somites of irregular shape and size (Bessho 2001). Loss of *Hes7* expression also results in the constant expression of *Lfng* throughout the PSM rather than the normal three distinct phases of expression (Bessho 2001). Additionally, sustained expression of *Hes7* results in fused somites (Takashima 2011). The severe abnormalities that result from any divergence from normal *Hes7* expression indicates that *Hes7* is a critical component of the segmentation clock.

In the mouse PSM, *Notch1* activates *Lfng* and *Hes7*. Lunatic fringe in turn inhibits *Notch1*, creating a feedback loop to synchronize oscillations. *Hes7* then inhibits *Lfng* and itself, relieving inhibition of *Notch1* and providing the core mechanism for oscillatory gene expression (Bessho 2003, Kageyama 2012). As previously mentioned, both loss of *Hes7* expression and sustained expression lead to somite fusion (Bessho 2001; Takashima 2011). The multiple layers of *Hes7* regulation and the severe phenotypes associated with *Hes7* perturbations indicate that *Hes7* is critically

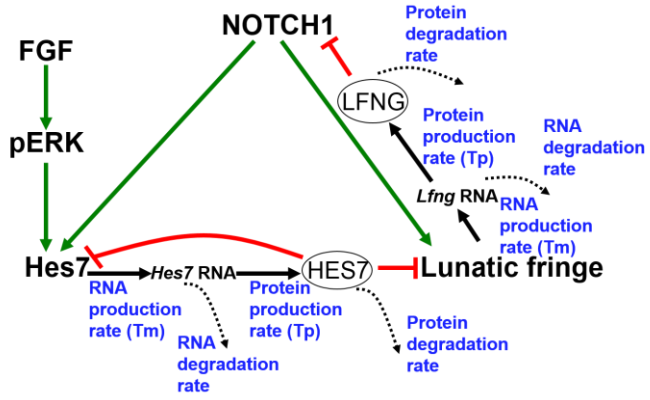

Figure 3. Segmentation clock regulation by feedback loops
Notch1 activates *Hes7* and *Lfng*. *Hes7* inactivates *Lfng* and itself. *Lfng* inactivates *Notch1*.

important to the segmentation clock. Furthermore, *Hes7* has been suggested to be the core pace-maker of the segmentation clock because *Hes7* initiates the 'off' phase of the clock by inactivating itself and *Lfng* (Kageyama 2012).

Hes7 and the clock

Hes7 expression exhibits two-hour oscillations across the PSM with three distinct phases similar to those of Lunatic fringe (Bessho 2001). The completion of one full oscillation, or one 'on/off' cycle, across the PSM results in the formation of one pair of somites. However, individual cells only register if they are 'on' or 'off' as shown in Figure 4, not whether they are in Phase I, II, or III as diagramed in Figure 1B. The PSM exhibits the phases of expression described in Figure 1 because variation in periodicity across the PSM causes the appearance of a wave of gene expression that moves anteriorly through the PSM rather than the entire PSM oscillating as a single entity. Therefore, the endogenous expression pattern of *Hes7* relies on the slowing oscillations in the anterior PSM but we do not know what is causing the oscillations to slow.

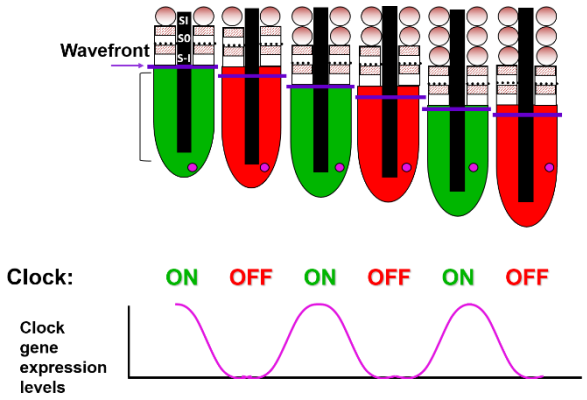

Figure 4. Clock Gene Expression
Cycles of clock gene expression stimulate the formation of new somites. Every on/off cycle produces one pair of somites.

Post-transcriptional clock regulation

In mathematical models of the clock, rapid oscillatory gene expression requires fast turnover of the mRNA and protein (Lewis 2003). Additionally, sustained oscillations, or

oscillations that can continue to repeat indefinitely with the same period and intensity, require two conditions be met: a short half-life of gene products and a long total delay time between gene activation and the formation of the completed protein (Kageyama 2012). This means that once the gene is turned on, it must do its job and then all evidence of its activation must quickly be removed in order to exhibit an 'off' phase. If transcription has ended but the mRNA is not degraded, then the gene expresses as 'on' for longer than intended. Therefore, oscillation periods are directly related to the rates of production and degradation of both the mRNA and protein. *Hes7* has been proposed to be responsible for setting the pace of the clock. If this is true, then the half-lives of *Hes7* mRNA and protein should be critical in determining the clock pace.

Mathematical models predict that protein half-life can influence the pace of the oscillations if they are within a certain period (Kageyama 2014; Feng 2007). The typical half-life of the HES7 protein is 22.3 minutes. Mutant mice were created to express a stabilized Hes7 protein with a half-life of 30.3 minutes (Hirata 2004). Oscillations were normal for four cycles before dampening and halting altogether. This dampened expression resulted in fused somites and severe skeletal defects, indicating that protein half-life directly influences oscillation. Experimental data supports the model where altering the HES7 protein half-life alters oscillation and breaks the clock, implying a direct link between HES7 protein half-life and the function of the clock.

Protein half-life is not the only potential factor in determining the period of oscillations. The rate of RNA production and degradation are also key players. Part of the delay in RNA production has ascribed to the process of splicing, or removing introns, from the immature RNA. *Hes7* has three introns, and the delay resulting from splicing has been measured as

approximately 19 minutes (Takashima 2011). Mathematical models indicate that stable *Hes7* oscillations may require this delay (Hirata 2004). When all three introns were deleted in mouse embryos to increase the speed of RNA maturation and decrease the total delay time, *Hes7* did not oscillate (Takashima 2011). This signals that the delays introduced by splicing introns are necessary for oscillatory expression. In theory, slightly extending or reducing the intronic delay should lengthen or accelerate oscillations. Deleting one or two introns was found to increase the number of cervical vertebrae, though mice with two deleted introns had more defects and fused vertebrae than mice with only one intron deletion (Harima 2013). This suggests that at least two introns are necessary for normal somitogenesis. When oscillation speed was examined in mutants, mutants with fewer introns were found to have more rapid oscillations and this accelerated clock resulted in the formation of extra cervical vertebra (Harima 2013). This supports the idea that shortening the intronic delay will accelerate the segmentation clock.

Many attempts have been made to slow the clock by extending intronic delay. When the Fujimuro group inserted an additional intron into the *Hes7* 3' untranslated region (3'UTR), the resulting transcript did not splice correctly and the endogenous 3'UTR was not included in the *Hes7* transcript. This resulted in a mouse with a skeleton resembling that of a *Hes7* null mutant. Other somite markers were irregularly expressed in the skeleton and the cyclic phases of *Hes7* expression in the PSM were lost (Fujimuro 2014). This indicates that the *Hes7* 3'UTR may be necessary for normal segmentation.

As mentioned previously, there are four components to setting the pace of oscillatory gene activity: mRNA production, protein production, mRNA degradation, and protein degradation. There is evidence that altering the rate of mRNA production on protein stability impacts the clock. Therefore, it is logical that mRNA degradation would also be critical to

determining the pace of the clock. Furthermore, altering the mRNA half-life as cells move into the anterior PSM could influence the slowing of the clock. This could be achieved 3'UTR-dependant regulation of transcript stability.

3'UTR and transcript half-life

The 3'UTR of a transcript is known to influence its mRNA stability and translational efficiency (Gupta 2014). Degradation of mRNA is primarily controlled through the 3'UTR because RNA-binding proteins (RBPs) and microRNAs (miRNAs) can bind to specific regions of the 3'UTR to destabilize the mRNA and lead to degradation or influence translation (Alonso 2012; Tian 2014). mRNAs bind to complementary sites on the mRNA and either induce degradation or repress translation (Xie 2007; Alonso 2012). Therefore, miRNAs can impact both mRNA stability and translational efficiency through binding sites on the 3'UTR. Evidence suggests that longer 3'UTRs are generally less stable than shorter 3'UTRs (Akman 2014; Davis 2014; Gupta 2014). This instability in long UTRs has been attributed in part to the increased availability of miRNA binding sites (Akman 2014; Davis 2014). Additionally, miRNA expression is believed to be highly tissue specific and heavily utilized to regulate mRNA degradation during development (Xie 2007). This could account for the highly regulated expression patterns and oscillations of the segmentation clock.

Alternative polyadenylation

Many genes are capable of producing transcript isoforms containing alternative 3'UTRs of varying length due to the presence of multiple polyadenylation (pA) sites within the gene region (Akman 2014; Davis 2014; Gupta 2014). This process is called alternative

polyadenylation (APA) and produces different isoforms of a single gene. In mice, the 3'UTRs of genes which utilize APA can vary in length from approximately 250 bp to 1,800 bp (Tian 2013). This means there may be a wide range of mRNA stability in these genes. Additionally, alternative polyadenylation allows additional levels of regulation through RBP and miRNAs. RBPs can both impact mRNA stability and influence the use of a pA site (Tian 2014). In mammalian genes that utilize APA, over half of the conserved miRNA target sites are in a long 3'UTR isoform (Tian 2014). Because long 3'UTRs have more miRNA or RBP binding sites, they potentially have more factors binding to the 3'UTR which can influence mRNA degradation or repress translation (Akman 2014; Tian 2014). This could be why messages with long 3'UTRs are frequently less stable than those with short 3'UTRs.

All previous research on *Hes7* in mice utilized a 3'UTR of approximately 265 bp (for example Nitanda 2014). However, we noticed that the genomic sequences downstream of the *Hes7* coding region are highly conserved between humans and mice and contain several potential polyadenylation sites that might, if utilized, give rise to three different mouse *Hes7* isoforms of different length (Fig. 5A). Previous work in our lab identified transcripts in mouse embryos using all three pA sites (Fig. 5B). Unpublished data from our lab demonstrates that isoforms with longer

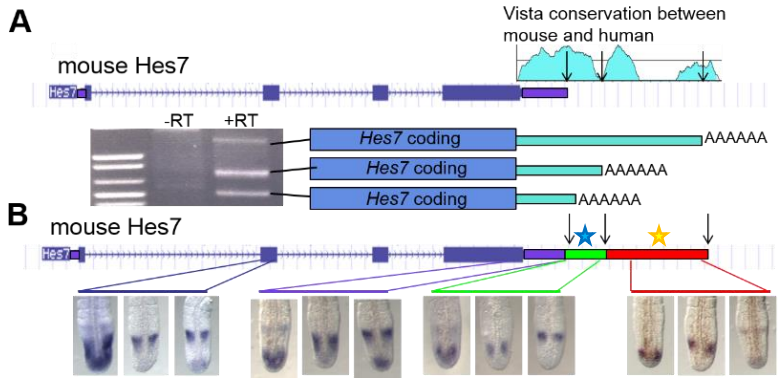

Figure 5. Alternative polyadenylation of mouse Hes7
A) The mouse *Hes7* locus with annotated UTRs in purple and Vista conservation of 3' regions, with arrows indicating conserved 3' regions.
B) Transcripts containing sequences corresponding to different 3'UTR regions oscillate in the mouse PSM. The blue star represents the mir346 binding site and the yellow star represents the mir99/100 binding site.

3'UTRS are depleted in the anterior region of the PSM where the clock period is slowest. This could indicate that the use of alternative pA sites may influence *Hes7* turnover in the PSM and contribute to the change in clock frequency across the PSM. Additionally, two different miRNA binding sites were located in the *Hes7* UTRs using TargetScan (www.targetscan.org). *Mir346* is found only in the middle-length and long UTRs, while *mir99/100* is found only in the long UTR, as shown in Fig. 5B. These miRNAs could be utilized to regulate *Hes7* mRNA half-life.

Research Justification

Preliminary data from our lab suggests that *Hes7* undergoes alternative polyadenylation to produce three different isoforms. Isoforms with longer 3'UTRs are depleted in the anterior PSM – the region with the slowest clock period. This suggests that the use of alternative polyadenylation may influence *Hes7* turnover in the PSM and contribute to the change in clock frequency across the PSM. Additionally, there are two different miRNA binding sites in the UTRs. It is attractive to speculate that alternative polyadenylation is utilized in the PSM to maximize the use of miRNAs to regulate the speed of mRNA turnover. This regulation in turn could control the pace of the segmentation clock through *Hes7*.

While it is an attractive theory, no research has been published on the topic. Additionally, we are left with several lasting questions. There is no information on whether 3'UTR isoforms have functional differences or if the miRNA binding sites are utilized. If they are utilized, how large of a role do they play in the degradation of mRNA? Finally, how is alternative polyadenylation regulated? Several gene-specific theories are available in the literature, but they have no experimental backing and there are no theories on *Hes7* specifically.

Ultimately, more research is necessary to determine how *Hes7* is regulated and why its UTR is necessary to somitogenesis.

My research is intended to address some of these issues. Specifically, I plan to examine the impact of APA on *Hes7* transcript half-life and stability. By forcing cells to produce only one of the three *Hes7* 3'UTR isoforms, I can examine the impact of the different length UTRs on transcript stability. I hypothesize that all *Hes7* UTRs will destabilize the transcript and the longer UTR isoforms will be more destabilizing than the short isoform. This will result in the long isoform producing a smaller quantity of a reporter protein than the short isoform, with the middle isoform producing an intermediate amount of protein.

CHAPTER 2: RESULTS

Site directed mutagenesis of *Hes7* 3'UTR

In order to analyze the differences between the 3'UTR isoforms, we first deleted the internal polyadenylation sites to create forced-expression isoforms as shown in Figure 6.

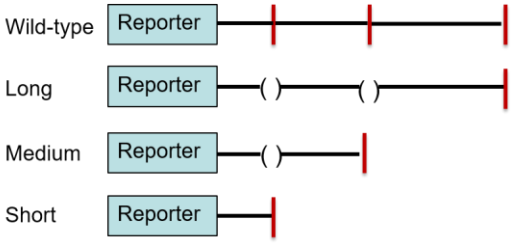

Figure 6. Visual representation of mutated 3'UTRs
Red lines denote intact polyadenylation sites. Parenthesis denote deleted segments.3'UTRs were later included in reporter constructs

We used fusion PCR (diagramed in Fig. 7) to delete the internal pA sites, TA cloned the mutated UTRs, and sequenced the resulting vectors to determine whether we successfully created the mutations. Figure 8 shows a sample of the sequencing results where the first pA site has been deleted from the medium forced-length UTR. These forced-length UTRs were inserted into pMIR vectors (Ambion commercial plasmid) containing the reporter protein Firefly Luciferase. Figure 9 is the image of an agarose gel where the individual forced-length UTR inserts were cut out of the pMIR backbone using

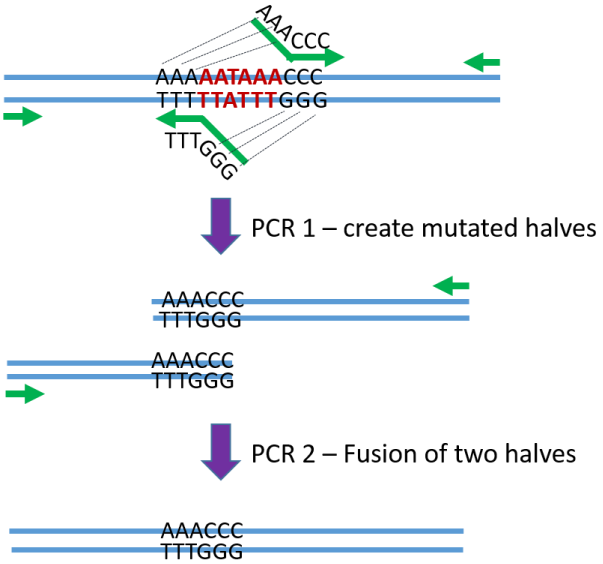

Figure 7. Fusion PCR diagram
Visual representation of the fusion PCR procedure.

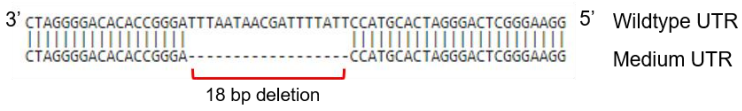

Figure 8. Sequencing
A portion of the wildtype UTR sequence is shown on top. The corresponding sequence from the middle UTR is on the bottom. The 18 base pair deletion, highlighted in red, shows where the 18 bases composing the first pA site have been deleted in the medium forced-length UTR.

restriction enzymes, demonstrating the successful insertion of the forced-length UTRs into pMIR.

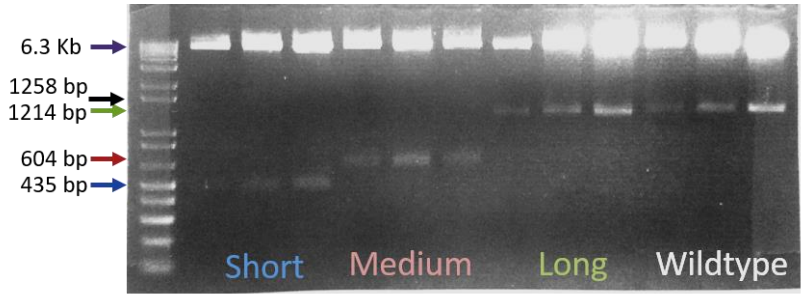

Figure 9. Digested Luciferase vectors Each vector composed of PMIR and a forced-length UTR was digested using XhoI and NotI three times. The top band, highlighted with a purple arrow, represents the PMIR backbone. The other bands, highlighted with blue, red, green, and black arrows, correspond to the short, medium, long, and wildtype UTRS respectively.

Confirmation of UTR utilization

We used 3'RACE to confirm that the deletion of the pA site will result in the forced expression of a single *Hes7* 3'UTR isoform. This technique, diagramed in Figure 10, creates cDNA from mRNA and utilizes a special oligo-dT adapter that will bind to the polyA tail of the mRNA and allow amplification of the entire 3' UTR. The cDNA is then PCR amplified using a gene-specific forward primer and a reverse primer that binds to the oligo-dT adapter sequence. When the PCR is run out on a gel, there will be bands specific to the size of the UTR utilized. I originally attempted to use transiently transfected cells for this experiment, but were unable to remove plasmid contamination to obtain clear results. Because of this, I have included data collected by Kara Braunreiter in which she utilized stable cells lines expressing Tet-inducible vectors containing the exogenous protein Venus and a forced-length *Hes7* 3'UTR. When we

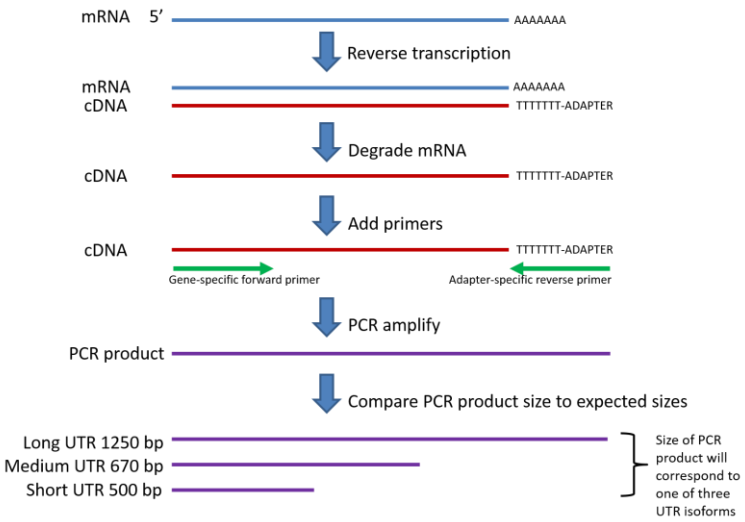

Figure 10. 3'RACE diagram Visual representation of the 3'RACE procedure.

collected RNA from these stable-expressing cell lines and performed 3'RACE, we found that each forced-length UTR gave rise to a single UTR while the wildtype UTR yielded three different isoforms (Fig. 11). This means our mutations are doing what they were intended to do: force the expression of a single isoform. However, there are additional bands present in our 3'RACE that are also present in a single-primer control sample. Because they are present in this control, we believe they are nonspecific bands and they do not indicate that our mutations were unsuccessful in forcing the expression of a single isoform.

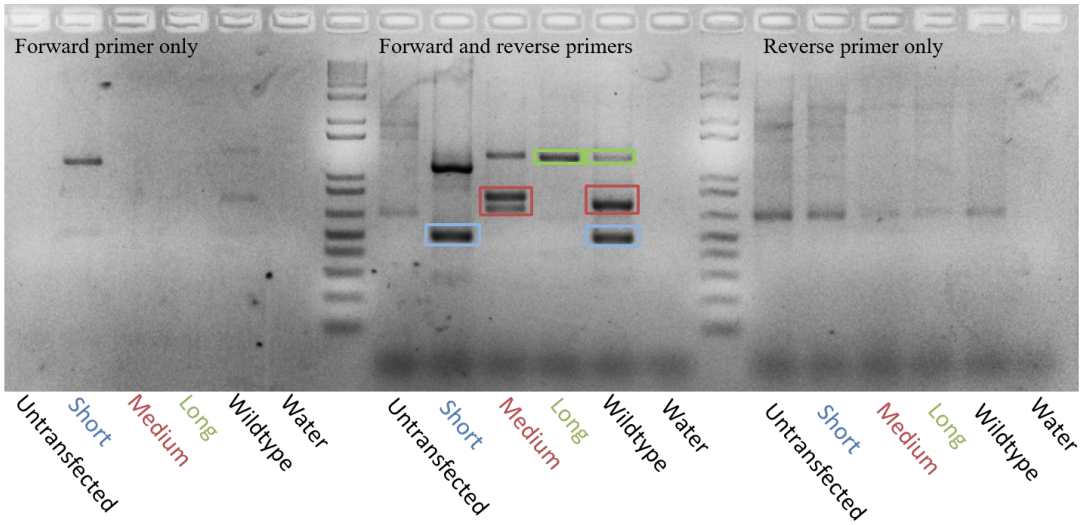

Figure 11. 3'RACE data
Bands of the correct size are highlighted and color-coded to the correct UTR isoform. Bands not highlighted are nonspecific bands also present in a control sample where only one primer was used. The small UTR is 500 bp, the medium UTR is 670 bp, and the long UTR is 1250 bp. Data shown was collected by Kara Braunreiter. Each forced-length UTR gave rise to a single UTR-sized band and the wildtype UTR produced all three isoforms. This indicates that our mutations do force the expression of a single isoform.

Examine effects of 3'UTR on Luciferase Expression

We wanted to see if the 3'UTR changes RNA stability or translational efficiency by comparing luciferase protein levels from the constructs created above. Plasmids encoding luciferase and the different mutated 3'UTRs were transfected into C2C12 cells and protein was collected and quantified using a luciferase assay. A change in protein level between the different

UTRs would indicate a change in RNA stability or translational efficiency. First, the wildtype *Hes7* 3'UTR, which can produce all three isoforms and therefore produces a mixture of all three transcripts, was compared with pMIR – the control vector containing no *Hes7* 3'UTR. Results shown in Figure 12A were normalized to Luciferase levels in pMIR. The addition of the wildtype 3'UTR reduced the amount of Luciferase protein present by 75% and this difference found to be statistically significant with a p-value of $p < 0.001$. This suggests that inclusion of the *Hes7* 3'UTR influences RNA stability and/or transcript translational efficiency.

To determine whether the individual 3'UTR isoforms have different effects, we compared the effects of the wildtype UTR to each forced-length isoforms. These results, shown in Figure 12B, were normalized to the wildtype 3'UTR. We found that the different UTR

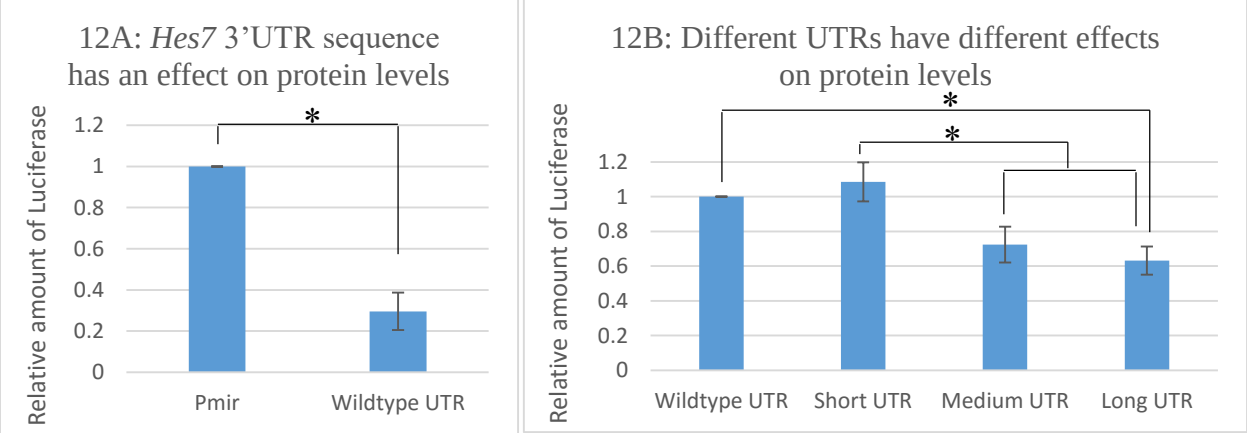

Figure 12. *Hes7* 3'UTR Luciferase Assays
12A) Vector with no *Hes7* 3'UTR compared to vectors containing the wildtype *Hes7* UTR. The wildtype UTR decreased the amount of protein. Error bars represent the standard error of the mean. * indicates $p < 0.001$
12B) Wildtype UTR compared to Short, Medium, and Long forced expression 3'UTRs. Error bars represent the standard error of the mean. * indicates $p < 0.05$

isoforms did produce different amounts of protein. As expected, the longer length UTRs produced less protein than the short UTR. The differences between the wildtype and long, short and medium, and short and long 3'UTRs were found to be statistically significant with a p-value

of $p < 0.05$. This suggests that the length of the 3'UTR is a determining factor of the stability or the translational efficiency of the mRNA.

Creation of miRNA site mutations

Two potential miRNA binding sites were identified in the *Hes7* 3'UTR. To determine whether they affect the stability of the different Hes7 isoforms, we deleted the miRNA binding sites using fusion PCR and created vectors for use in future experiments (Fig. 13A). Figure 13B shows the individual fragments utilized in the PCR and Figure 13C shows the fused medium346 and long99/100 segments later inserted into the PMIR backbone. Currently, no experiments have been performed utilizing these UTRs.

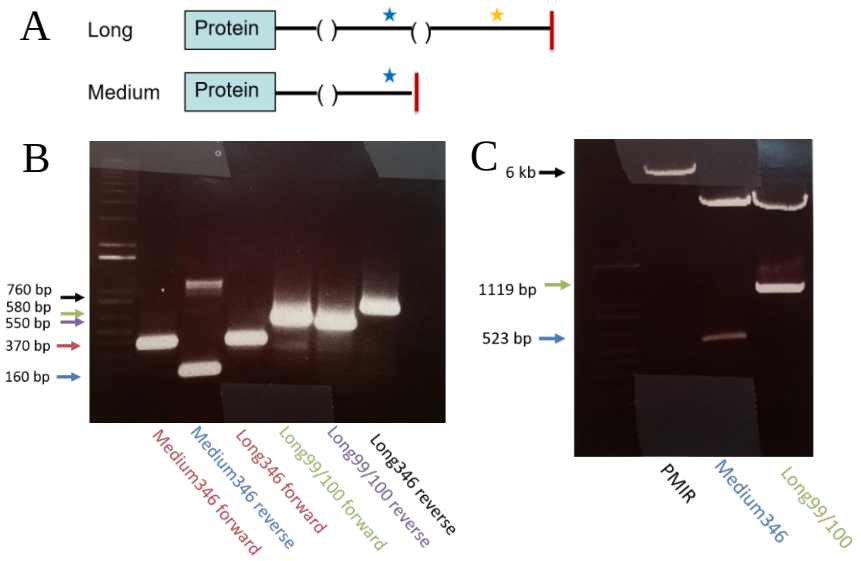

Figure 13. Mutated miRNA site vectors
13A) Diagram of the mutated miRNA vectors. The mir346 site is represented with a blue star and the mir99/100 site is represented with a yellow star.
13B) Gel showing each fragment utilized for the fusion PCR. These bands were gel extracted and fused to form the completed mutated miRNA UTRs.
13C) Each mutated miRNA UTR has been fused and TA cloned. Here, the UTRs were extracted using MluI and XbaI.

CHAPTER 3: METHODS

Site directed mutagenesis of *Hes7* 3'UTR

PCR mutagenesis

The short and medium length UTRs were amplified from genomic DNA using fusion PCR (diagramed in Fig. 7) and the following primers with added restriction enzyme cut sites italicized:

A: *Hes7* forward with added *XbaI* cut site: *TATCTAGA*TTGGCCCTGAGCTTTTGG

B: Short UTR reverse with added *ClaI* cut site: *TAATCGAT*GGAGAAGGGTATGGGCAAGT

C: Medium UTR reverse with added *ClaI* cut site: *TAATCGAT*GCGATTCAAAGGTTGTGG

D: pA1 deletion forward: CGAGTCCCTAGTGCATGGTCCCGGTGTGTCCCCTA

E: pA1 deletion reverse: GCACTAGGGACTCGGGAAG

The short UTR was made using primers A and B. The first half of the medium UTR was made using primers A and E. The second half was made with primers C and D. The two pieces were ligated with A and C.

The long UTR was made using fusion PCR from the following primers with added restriction enzyme cut sites italicized:

A: *Hes7* forward with added *XbaI* cut site: *TATCTAGA*TTGGCCCTGAGCTTTTGG

F: Medium UTR reverse with *HindIII* cut site: *TAAAGCTT*CTTGGCTTTTAAAAAATTTAAAG

G: pA2 deletion forward with *HindIII* cut site: *TAAAGCTT*ACCGACTCTGGTGGGATTC

H: Long UTR reverse with added *ClaI* cut site: *TAATCGAT*TTCAAACCCGAAATGAGGAC

Primers A and F were used to amplify the medium UTR and add a *HindIII* cut site over the second pA signal (called Long AB). Primers G and H were used to amplify the region after the

second pA signal (termed Long C) and add a *HindIII* cut site over the pA signal from genomic DNA.

All PCR products were loaded onto agarose gels and bands of the appropriate size were gel extracted.

TA cloning and sequencing

The gel-extracted short, medium, long part AB, and long part C UTRs were TA cloned into TOPO vectors using an Invitrogen TOPO TA Cloning Kit with Dual Promoter pCR-II TOPO Vector and following the manufacturer's instructions. These vectors were sequenced to confirm that the mutations were successful.

Long AB was extracted from the TOPO vector using XbaI and HindIII. Long C was extracted with HindIII and ClaI. The two pieces were ligated over the common HindIII site and the resulting complete UTR was TA cloned and sequenced.

Construction of Luciferase vectors

The short, medium, long, and wildtype UTRs were extracted from TOPO vectors using *SpeI* and *NotI* and ligated into the Ambion commercial plasmid pMIR-REPORT (abbreviated pMIR) which contains the reporter protein Firefly Luciferase.

Confirmation of UTR utilization

3'RACE; experiment performed by Kara Braunreiter

3'RACE (diagramed in Fig. 10) was utilized to confirm that the deleted polyadenylation sites was sufficient to force the expression of a single 3'UTR. The 3'UTRs described above were inserted into a vector with a Tet-inducible promoter and Venus protein. Stable cell lines expressing these vectors were produced by Kara Braunreiter. MRNA was Trizol-extracted from these cell lines and used to reverse-transcribe cDNA using an Invitrogen SuperScript-III first strand kit. The cDNA was PCR amplified using a forward primer that binds to the Venus coding-region and the reverse Gene-racer 3' Primer from the SuperScript-III kit. The PCR product was run on an agarose gel and band sizes were used to identify the UTRs.

Examine effects of 3'UTR on Luciferase Expression

Luciferase analysis

The Luciferase-UTR constructs were transfected into C2C12 cells plated in 24-well plates containing 20,000 cells. 450ng of the Luciferase vector and 50ng of Renilla, a pRL vector containing Renilla luciferase as a control reporter vector, were co-transfected using Lipofectamine 2000 from Invitrogen, following the manufacturer's instructions. Renilla served as an internal standard to normalize transfection efficiency. After approximately 48 hours of transfection, cells were lysed and protein content was measured via a luminosity test using a Promega Dual-Luciferase assay system kit, following the manufacturer's directions. This experiment was repeated three times in triplicate.

Statistical analysis

The luminosity value of Luciferase was divided by the luminosity value of the renilla to give a luciferase:renilla ratio (L/R) for each well of cells. This normalized the transfection efficiency between each replicate. The three L/R values per UTR were averaged and normalized to either PMIR or the wildtype UTR. These normalized values were then averaged across three replicates to give one final protein quantification shown in Figure 7. All results were analyzed using an ANOVA test followed by a Student Newman Keuls posthoc test to determine statistical significance with a p-value of at least $p < 0.05$.

Creation of miRNA site mutations

PCR mutagenesis

The miRNA sites were deleted using fusion PCR, similar to the deletion of the pA sites previously. All miRNA sites were replaced with HindIII cutsites (underlined in sequences below). The mir346 site was deleted from the middle and long forced-length UTRs using PCR and the following primers with restriction enzyme cut sites italicized:

A: *Hes7* forward with added *XbaI* cut site: *TATCTAGA*TTGGCCCTGAGCTTTTGG

I: *Hes7*mir346 replaced with *HindIII* reverse: T*AAGCTT*CCAGGGCTTG

J: Medium UTR reverse with added *MluI* cut site: *TAACGCGT*GCGATTCAAAGGTTGTGG

K: *Hes7*mir346 replaced with *HindIII* forward: TGG*AAGCTT*AGGTAAGTAAA

L: Long UTR reverse with added *MluI* cut site: *TAACGCGT*TCAAACCCGAAATGAGGAC

The medium UTR with deleted mir346 site (called medium346) was made in two parts using the medium forced-length UTR as a template. Part 1 was made using primers A and I, and part 2 was made using primers J and K. Parts 1 and 2 were fused using primers A and J. The long

UTR with deleted mir346 site (called long346) was also made in two parts using the long forced-length UTR as a template. Part 1 was made using primers A and I, and part 2 was made using primers L and K. Part 1 and 2 were fused using primers A and L.

The long UTR with deleted mir99/100 site (called long99/100) was made using fusion PCR from the following primers:

A: *Hes7* forward with added *XbaI* cut site: TATCTAGATTGGCCCTGAGCTTTTGG

L: Long UTR reverse with added *MluI* cut site: TAACGCGTTCAAACCCGAAATGAGGAC

M: *Hes7*mir99/100 replaced with *HindIII* reverse: GAAGCTTAGCCGGAAG

N: *Hes7*mir99/100 replaced with *HindIII* forward: GCTAAGCTTCCAGTATAGGAAA

Long99/100 was made in two parts using the long forced-length UTR as a template. Part 1 was made using primers A and M, and part 2 was made using primers L and N. Part 1 and 2 were fused using primers A and L.

All PCR products were loaded onto agarose gels and bands of the appropriate size were gel extracted.

TA cloning and sequencing

The gel-extracted medium346, long346, and long99/100 UTRs were TA cloned into TOPO vectors as above and sequenced to confirm that the mutations were successful.

Construction of Luciferase vectors

Medium346 and long99/100 were extracted from the TOPO vector using a sequential digest of *MluI* followed by *XbaI*. The PMIR backbone was linearizes using *MluI* and *SpeI*. Because *SpeI* and *XbaI* are compatible, the UTRs were able to be ligated into the backbone containing the reporter protein Firefly Luciferase.

CHAPTER 4: DISCUSSION AND CONCLUSIONS

During the process of somitogenesis, members of the Notch signaling pathway exhibit oscillatory gene expression. The period of these oscillations slows in the anterior PSM, but the mechanism for this is unknown. *Hes7* has been suggested to be the core pace-setter of the molecular clock and therefore any mechanisms regulating *Hes7* oscillations are important to understand in terms of identifying factors which influence the period of the entire clock. Here, we propose that alternative polyadenylation contributes to regulation of the *Hes7* 3'UTR to alter its oscillation speed, and therefore the oscillation period of the entire clock, across the PSM. My research examined the three different *Hes7* 3'UTR isoforms to determine whether there are any functional differences between the isoforms that would support our potential model.

To test our model, we deleted the polyadenylation signals from the *Hes7* 3'UTR to create what we have called forced-length UTRs. The effects of these mutated UTRs on a reporter protein were measured using Luciferase assays. We expected any *Hes7* 3'UTR to cause a decrease in protein because we believe the *Hes7* 3'UTR has a destabilizing effect on the mRNA to cause rapid degradation of the RNA and maintain rapid oscillations. We also expected the UTR isoforms to have different effects such that the short UTR produces the most protein and the long UTR produces the least with the medium UTR falling somewhere between them because long UTRs are typically less stable than shorter UTRs and an intermediate length UTR should therefore have an intermediary stability. We found that the presence of any *Hes7* 3'UTR causes a significant decrease in the amount of protein produced compared to a vector without a *Hes7* 3'UTR. Additionally, the long and medium UTRs both produce less protein than the short UTR. These results show that there are functional differences between the different UTR

isoforms. We believe these differences could be a mechanism to adjust the speed of *Hes7* oscillations across the PSM.

While we have shown that the *Hes7* 3'UTR decreases the amount of protein present in a luciferase reporter assay, we do not know how it does so. 3'UTRs can play roles in both mRNA half-life and translational efficiency. A decrease in either of these factors would result in a decrease in the amount of protein produced in the assays we utilized. Decreasing the stability of the mRNA causes a decrease in the amount of steady state RNA present in the cell because the RNA has a shorter half-life and degrades faster. This causes a decrease in protein because there is less template available for translation. Alternatively, the 3'UTR and polyA tail assist with initiating translation and therefore impact translational efficiency. Even if steady state mRNA levels are identical, different amounts of protein could be produced because one mRNA is harder or slower to translate than another.

In summary, we have two potential mechanisms that both would result in decreased protein levels. Potential future directions would determine whether RNA stability or translational efficiency is responsible for causing the decrease in protein levels. RNA stability can be measured using stable cell lines and qPCR. A construct containing a Tet-inducible promoter, exogenous fluorescing Venus protein, and forced-length UTR would be used to create stable-expressing cell lines. Addition of doxycycline or tetracycline to the media would halt transcription by inhibiting initiation from the Tet-promoter so no new RNA could be made. RNA would be collected over time and the quantity of mRNA at set time-points would be measured using qPCR. The resulting degradation curve would allow us to deduce the mRNA half-life. If the *Hes7* 3'UTR is altering mRNA stability, then we would expect transcripts containing the short UTR to have a longer half-life than transcripts containing the long UTR and

our results could be explained by the fact that a longer transcript half-life would result in the production of more protein. If the half-lives are similar between the different isoforms, then the differences in protein levels are likely due to translation efficiency.

Translational efficiency can be determined by measuring steady state RNA and protein levels. Cells transfected with the Luciferase vectors would be harvested at the same time and half of the harvested cells would be used to measure protein levels via a Luciferase assay. RNA would be collected from the remaining cells and quantified using qPCR. The ratio of RNA to protein can then be used to calculate translational efficiency. If the 3'UTRs are impacting translational efficiency, then we would predict the short UTR to have a higher ratio of protein to RNA than the long UTR, explaining how the short UTR can make more protein per RNA molecule than the long UTR can. If translational efficiency is not regulated by the 3'UTR, then all the isoforms will have similar protein to RNA ratios.

A third future direction would attempt to identify if the variation between the individual UTRs is due to miRNA activity. MiRNAs can be co-transfected with Luciferase vectors into cells that do not already contain the miRNA in question and a luminosity assay would measure the amount of protein present. Protein levels would be compared between vectors with intact miRNA binding sites and vectors with deleted miRNA binding sites. If the miRNA targets the transcript, then we would expect to see lower protein levels in the UTRs with intact binding sites because the miRNA would destabilize the mRNA or inhibit translation and therefore cause a lower protein yield than if the miRNA does not bind to the mRNA. If we see that miRNAs are binding to the mRNA, we could then perform in situ hybridization on mouse embryos to analyze the expression patterns of the miRNAs that target *Hes7* in the PSM.

A final experiment would be to create transgenic mice. These mice would contain a vector containing the *Hes7* promoter, sequences encoding the exogenous Venus protein, and a forced-length UTR. Ideally, we would create a line of mice for each of the three *Hes7* 3'UTR isoforms. We could then collect embryos and use in situ hybridization to examine Venus transcript expression. The promoter and UTR would control the location of Venus expression in the PSM and amount of protein present in the mouse, so any difference in expression patterns between mice containing different UTR isoforms would be dictated by the UTR itself. This would allow us to see exactly how each UTR isoform influences expression in the.

Though the use of alternative polyadenylation is a plausible mechanism for altering the speed of the clock across the PSM, we currently do not know how a specific UTR isoform is selected for in varying regions of the PSM. There may be spatial cues in the PSM, such as the Wnt or FGF expression gradients, which influence the usage of one pA site over another. In mice, FGF oscillations specify the location of the wavefront, Notch signaling regulates the clock, and *Hes7* links the two (Niwa 2011). Therefore, we know that FGF can influence *Hes7* expression to coordinate spatial and temporal synchrony in the PSM. If FGF can influence the preferential use of pA sites, then spatial information can be incorporated into the pace of the clock via *Hes7* so that the anterior PSM experiences slower oscillations than the posterior PSM.

We have proposed that alternative polyadenylation of the *Hes7* 3'UTR is responsible for altering the pace of the clock across the PSM. By preferentially expressing the short, more-stable UTR where the clock has a longer period and the long, less-stable UTR where the clock period is shorter, the entire pace of the clock can be specified to location. This has evolutionary significance because the process of somitogenesis and the clock genes themselves are highly conserved among vertebrates. Both zebrafish and mice are known to have a slowing clock

period, and it is not a far reach to speculate that this is also a common feature in vertebrate development. Therefore, this mechanism of using alternative polyadenylation to adjust oscillation periods might also be evolutionarily conserved and should be investigated in organisms and other systems.

LITERATURE CITED

Akman, H., & Erson-Bensan, A. (2014). Alternative polyadenylation and its impact on cellular processes. MicroRNA, 3(1), 2-9.

Alonso, C. R. (2012). A complex 'mRNA degradation code' controls gene expression during animal development. Trends in Genetics, 28(2), 78-88.

Bessho, Y., Sakata, R., Komatsu, S., Shiota, K., Yamada, S., & Kageyama, R. (2001). Dynamic expression and essential functions of Hes7 in somite segmentation. Genes & Development, 15(20), 2642-2647.

Bessho, Y., Hirata, H., Masamizu, Y., Kageyama, R. (2003). Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. Genes & Development, 17, 1451-1456.

Cooke, J., & Zeeman, E. C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. Journal of Theoretical Biology, 58(2), 455-476.

Davis, R., & Shi, Y. (2014). The polyadenylation code: A unified model for the regulation of mRNA alternative polyadenylation. Journal of Zhejiang University.Science.B, 15(5), 429-437.

Feng, P., & Navaratna, M. (2007). Modelling periodic oscillations during somitogenesis.. Mathematical Biosciences and Engineering, 4(4), 661-673.

Ferjentsik, Z., Hayashi, S., Dale, J. K., Bessho, Y., Herreman, A., De Strooper, B., et al. (2009). Notch is a critical component of the mouse somitogenesis oscillator and is essential for the formation of the somites. PLoS Genetics, 5(9), e1000662.

Fujimuro, T., Matsui, T., Nitanda, Y., Matta, T., Sakumura, Y., Saito, M., et al. (2014). Hes7 3'UTR is required for somite segmentation function. Scientific Reports, 4, 6462.

González, A., Manosalva, I., Liu, T., & Kageyama, R. (2013). Control of Hes7 Expression by Tbx6, the Wnt Pathway and the Chemical Gsk3 Inhibitor LiCl in the Mouse Segmentation Clock. PLoS ONE, 8(1), e53323.

Gupta I, Clauder-Münster S, Klaus B, Järvelin AI, Aiyar RS, Benes V, Wilkening S, Huber W, Pelechano V, Steinmetz LM. (2014). Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. Molecular Systems Biology, 10(2), 719.

Harima, Y., Imayoshi, I., Shimojo, H., Kobayashi, T., & Kageyama, R. (2014). The roles and mechanism of ultradian oscillatory expression of the mouse hes genes. Seminars in Cell & Developmental Biology, 34, 85-90.

Harima, Y., Takashima, Y., Ueda, Y., Ohtsuka, T., & Kageyama, R. (2013). Accelerating the tempo of the segmentation clock by reducing the number of introns in the Hes7 gene. Cell Reports, 3(1), 1-7.

Hirata, H., Bessho, Y., Kokubu, H., Masamizu, Y., Yamada, S., Lewis, J., et al. (2004). Instability of Hes7 protein is crucial for the somite segmentation clock. Nature Genetics, 36(7), 750-754.

Kageyama, R., Niwa, Y., Isomura, A., González, A., & Harima, Y. (2012). Oscillatory gene expression and somitogenesis. Wiley Interdisciplinary Reviews: Developmental Biology, 1(5), 629-641.

Lewis, J. (2003). Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. Current Biology, 13(16), 1398-1408.

Nitanda, Y., Matsui, T., Matta, T., Higami, A., Kohno, K., Nakahata, Y., et al. (2014). 3′–UTR-dependent regulation of mRNA turnover is critical for differential distribution patterns of cyclic gene mRNAs. FEBS Journal, 281(1), 146-156.

Niwa, Y.,Masamizu, Y., Liu, T., Nakayama, R., Deng, C.-X, Kageyama, R. (2007) The initiation and propagation of Hes7 oscillation are cooperatively regulated by Fgf and Notch signaling in the somite segmentation clock. Developmental Cell, 13(2), 298-304.

Niwa Y,. Shimojo, H., Isomura, A., Gonzalez, A., Miyachi, H., Kageyama, R. (2011) Different types of oscillations in Notch and Fgf signaling regulate the spatiotemporal periodicity of somitogenesis. Genes and Development, 25, 1115-1120.

Penton, A. L., Leonard, L. D., & Spinner, N. B. (2012). Notch signaling in human development and disease. Seminars in Cell & Developmental Biology, 23(4), 450-457.

Pourquié, O. (2011). Vertebrate segmentation: From cyclic gene networks to scoliosis. Cell, 145(5), 650-663.

Riley, M. F., Bochter, M. S., Wahi, K., Nuovo, G. J., & Cole, S. E. (2013). Mir-125a-5p-mediated regulation of lfng is essential for the avian segmentation clock. Developmental Cell, 24(5), 554-561.

Takashima, Y., Ohtsuka, T., González, A., Miyachi, H., & Kageyama, R. (2011). Intronic delay is essential for oscillatory expression in the segmentation clock. Proceedings of the National Academy of Sciences, 108(8), 3300-3305.

Tian, B., & Manley, J. L. (2013). Alternative cleavage and polyadenylation: The long and short of it. Trends in Biochemical Sciences, 38(6), 312-320.

Wahi, K., Bochter, M. S., & Cole, S. E. (2016). The many roles of notch signaling during vertebrate somitogenesis. Seminars in Cell & Developmental Biology, 49, 68-75.

Xie, Z., Yang, H., Liu, W., & Hwang, M. (2007). The role of microRNA in the delayed negative feedback regulation of gene expression. Biochemical and Biophysical Research Communications, 358(3), 722-726.